

```

1  TCCGGGGTCC GCACCGGGCC TGAGTCGGTC CGAGGCCGTC CCAGGAGCAG
51 CTGCCCCGTC GGAACAGCAC TATGGGCTTC TCTTCTGAGC TGTGCAGCCC
101 CCAGGGCCAC GGGGTCCTGC AGCAAATGCA GGAGGCCGAG CTTCGTCTAC
151 TGGAGGGCAT GAGAAAGTGG ATGGCCCAGC GGGTCAAGAG TGACAGGGAG
201 TATGCAGGAC TGCTTCACCA CATGTCCCTG CAGGACAGTG GGGGCCAGAG
251 CCGGGCCATC AGCCCTGACA GCCCCATCAG TCAGTCCTGG GCTGAGATCA
301 CCAGCCAAAC TGAGGGCCTG AGCCGCTTGC TGCGGCAGCA CGCAGAGGAT
351 CTGAACTCAG GGCCCTGAG CAAGCTGAGC CTGCTCATCC GGGAACGGCA
401 GCAGCTTCGC AAGACCTACA GCGAGCAGTG GCAGCAGCTG CAGCAGGAGC
451 TCACCAAGAC CCACAGCCAG GACATTGAGA AGCTGAAGAG CCAGTACCGA
501 GCTCTGGCAC GGGACAGTGC CCAAGCCAAG CGCAAGTACC AGGAGGCCAG
551 CAAAGACAAG GACCGTGACA AGGCCAAGGA CAAGTATGTG CGCAGCCTGT
601 GGAAGCTCTT TGCTCACCAC AACCGCTATG TGCTGGGCGT GCGGGCTGCG
651 CAGCTACACC ACCAGACCA CCACCAGCTC CTGCTGCCCC GCCTGCTGCG
701 GTCACTGCAG GACCTGCACG AGGAGATGGC TTGCATCCTG AAGGAGATCC
751 TGCAGGAATA CCTGGAGATT AGCAGCCTGG TGCAGGATGA GGTGGTGGCC
801 ATTACCCGGG AGATGGCTGC AGCTGCTGCC CGCATCCAGC CTGAGGCTGA
851 GTACCAAGGC TTCCTGCGAC AGTATGGGTC CGCACCTGAC GTCCCACCCT
901 GTGTCACGTT CGATGAGTCA CTGCTTGAGG AGGGTGAACC GCTGGAGCCT
951 GGGGAGCTCC AGCTGAACGA GCTGACTGTG GAGAGCGTGC AGCACACGCT
1001 GACCTCAGTG ACAGATGAGC TGGCTGTGGC CACCGAGATG GTGTTTCAGGC
1051 GGCAGGAGAT GGTACGCAG CTGCAACAGG AGCTCCGGAA TGAAGAGGAG
1101 AACACCCACC CCCGGGAGCG GGTGCAGCTG CTGGGCAAGA GGCAAGTGCT
1151 GCAAGAAGCA CTGCAGGGGC TGCAGGTAGC GCTGTGCAGC CAGGCCAAGC
1201 TGCAGGCCCA GCAGGAGTTG CTGCAGACCA AGCTGGAGCA CCTGGGCCCC
1251 GGCGAGCCCC CGCCTGTGCT GCTCCTGCAG GATGACCGCC ACTCCACGTC
1301 GTCCTCGGAG CAGGAGCGAG AGGGGGGAAG GACACCCACG CTGGAGATCC
1351 TTAAGAGCCA CATCTCAGGA ATCTTCCGCC CCAAGTTCTC GAACCTGTAC
1401 CGACTGGAAG GGAAGGCTT TCCTAGCATT CTTTGTCTCA TCGACCACCT
1451 ACTGAGCACC CAGCAGCCCC TCACCAAGAA GAGTGGTGTT GTCCTGCACA
1501 GGGCTGTGCC CAAGGACAAG TGGGTGCTGA ACCATGAGGA CCTGGTGTG
1551 GGTGAGCAGA TTGGACGGGG GAACTTTGGC GAAGTGTTCA GCGGACGCCT
1601 GCGAGCCGAC AACACCCTGG TGGCGGTGAA GTCTTGTCGA GAGACGCTCC
1651 CACCTGACCT CAAGGCCAAG TTTCTACAGG AAGCGAGGAT CCTGAAGCAG
1701 TACAGCCACC CCAACATCGT GCGTCTCATT GGTGTCTGCA CCCAGAAGCA
1751 GCCCATCTAC ATCGTCATGG AGCTTGTCGA GGGGGGCGAC TTCCTGACCT
1801 TCCTCCGCAC GGAGGGGGCC CGCCTGCGGG TGAAGACTCT GCTGCAGATG
1851 GTGGGGGATG CAGCTGCTGG CATGGAGTAC CTGGAGAGCA AGTGCTGCAT
1901 CCACCGGGAC CTGGCTGCTC GGAAGTGCCT GGTGACAGAG AAGAATGTCC
1951 TGAAGATCAG TGACTTTGGG ATGTCCCGAG AGGAAGCCGA TGGGGTCTAT
2001 GCAGCCTCAG GGGGCCTCAG ACAAGTCCCC GTGAAGTGGA CCGCACCTGA
2051 GGCCCTTAAC TACGGCCGCT ACTCCTCCGA AAGCGACGTG TGGAGCTTTG
2101 GCATCTTGCT CTGGGAGACC TTCAGCCTGG GGGCCTCCCC CTATCCCAAC
2151 CTCAGCAATC AGCAGACACG GGAGTTTGTG GAGAAGGGGG GCCGTCTGCC
2201 CTGCCCAGAG CTGTGTCCTG ATGCCGTGTT CAGGCTCATG GAGCAGTGCT
2251 GGGCCTATGA GCCTGGGCAG CGGCCAGCT TCAGCACCAT CTACCAGGAG

```

FIG. 1A

2301 CTGCAGAGCA TCCGAAAGCG GCATCGGTGA GGCTGGGACC CCCTTCTCAA
2351 GCTGGTGGCC TCTGCAGGCC TAGGTGCAGC TCCTCAGCGG CTCCAGCTCA
2401 TATGCTGACA GCTCTTCACA GTCCTGGACT CCTGCCACCA GCATCCACAC
2451 TGCCGGCAGG ATGCAGCGCC GTGTCTCTC TGTGTCCCTG CTGCTGCCAG
2501 GGCTTCCTCT TCCGGGCAGA AACAAATAAAA CCACTTGTGC CCACTGAAAA
2551 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2601 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2651 AAAAAAAAAA AAAAAAAAAA AAAA (SEQ ID NO:1)

FEATURES:

5'UTR: 1-71
Start Codon: 72
Stop Codon: 2328
3'UTR: 2331

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 18000004928524 /altid=gi 4503687 /def=ref NP_001996.1 feli...	1364	0.0
CRA 18000004928925 /altid=gi 400127 /def=sp P07332 FES_HUMAN PR...	1361	0.0
CRA 335001098689057 /altid=gi 11433086 /def=ref XP_007718.1 fe...	1361	0.0
CRA 18000004944482 /altid=gi 66835 /def=pir TVCTFF protein-tyr...	1285	0.0
CRA 18000004883448 /altid=gi 1345986 /def=sp P14238 FES_FELCA P...	1285	0.0
CRA 18000004938794 /altid=gi 125356 /def=sp P16879 FES_MOUSE PR...	1249	0.0
CRA 18000004944484 /altid=gi 125354 /def=sp P00543 FES_FSVST TY...	647	0.0
CRA 18000004958077 /altid=gi 323873 /def=gb AAA43041.1 (J02087...	621	e-176
CRA 18000004944483 /altid=gi 125353 /def=sp P00542 FES_FSVGA TY...	621	e-176
CRA 108000000500738 /altid=gi 7548235 /def=gb AAA43046.2 (J020...	603	e-171

BLAST dbEST hits:

	Score	E
gi 12875454 /dataset=dbest /taxon=960...	1669	0.0
gi 12259598 /dataset=dbest /taxon=960...	924	0.0
gi 5526793 /dataset=dbest /taxon=9606 ...	856	0.0
gi 1501859 /dataset=dbest /taxon=9606 ...	722	0.0
gi 9097978 /dataset=dbest /taxon=9606...	714	0.0
gi 6131861 /dataset=dbest /taxon=9606 ...	682	0.0
gi 12447497 /dataset=dbest /taxon=96...	674	0.0

FIG. 1B

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

From BLAST dbEST hits:

gi|12875454 Placenta

gi|12259598 Lung-tumor

gi|5526793 Kidney 2 pooled tumors

gi|1501859 Pregnant uterus

gi|9097978 Pediatric ore-B cell lymphoblastic leukemia

gi|6131861 Stomach poorly differentiated adenocarcinoma with signet ring cells

gi|12447497 Placenta normal

From tissue screening panels:

Hippocampus

1 MGFSSSEL CSP QGHGVLQMQ EAELRLLEG M RKWMAQRVKS DREYAGLLHH
51 MSLODGGQS RAISPDSPIS QSWAETTSQT EGLSRLLRQH AEDLNSGPLS
101 KLSLLIRERQ QLRKTYSEQW QQLQQELTKT HSQDIEKLKS QYRALARDSA
151 QAKRKYQEAS KDKDRDKAKD KYVRS LWKLF AHNRYVLGV RAAQLHHQH
201 HQLLLPGLLR SLQDLHEEMA CILKEILQEY LEISLVQDE VVAIHREMAA
251 AAARIQPEAE YQGFLRQYGS APDVPPCVTF DESLLEEGER LEPGELQLNE
301 LTVESVQHTL TSVTDELAVA TEMVFRQEM VTQLQQELRN EEENTHPRER
351 VQLLGKRQVL QEALQGLQVA LCSQAKLQAA QELLQTKLEH LGPGEP PPVL
401 LLQDDRHS TS SSEQEREGR TPTLEILKSH ISGIFRPKFS NLYRLEGE GF
451 PSIPLLIDHL LSTQQPLTKK SGVWLHRAVP KDKWLNHED LVLGEQIGRG
501 NFGEVFSGRL RADNTLVAVK SCRETLPPDL KAKFLQEARI LKQYSHPNIV
551 RLIGVCTQKQ PIYIMELVQ GGDFTFLRT EGARLRVKT LQMVGDAAG
601 MEYLESKCCI HRDLAARNCL VTEKNVLKIS DFGMSREEAD GUYAASGGLR
651 QVPVKWTAPE ALNYGRYSSE SDVWSFGILL WETFSLGASP YPNLSNQQTR
701 EFVEKGGRLP CPELCPDAVF RLMEQCWAYE PGQRPSFSTI YQELQSIRKR
751 HR (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

693-696 NLSN

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 7

1	40-42	SDR
2	468-470	TKK
3	507-509	SGR
4	521-523	SCR
5	557-559	TQK
6	622-624	TEK
7	746-748	SIR

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

FIG. 2A

Number of matches: 17

1	40-43	SDRE
2	52-55	SLQD
3	211-214	SLQD
4	72-75	SWAE
5	78-81	SQTE
6	115-118	TYSE
7	52-55	SLQD
8	211-214	SLQD
9	270-273	SAPD
10	279-282	TFDE
11	283-286	SLLE
12	312-315	SVTD
13	386-389	TKLE
14	410-413	SSSE
15	412-415	SEQE
16	521-524	SCRE
17	635-638	SREE

[4] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site

Number of matches: 3

1	37-44	RVKSDREY
2	254-261	RIQPEAEY
3	636-643	REEADGVY

[5] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 6

1	58-63	GQSRAI
2	189-194	GVRAAQ
3	207-212	GLLRSL
4	366-371	GLQVAL
5	641-646	GVYAAS
6	732-737	GQRPSF

[6] PDOC00009 PS00009 AMIDATION
Amidation site

354-357 LGKR

[7] PDOC00100 PS00107 PROTEIN_KINASE_ATP
Protein kinases ATP-binding region signature

497-520 IGRGNFGEVFSGRRLRADNTLVAVK

[8] PDOC00100 PS00109 PROTEIN_KINASE_TYR
Tyrosine protein kinases specific active-site signature

609-621 CIHRDLAARNCLV

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	672	692	1.129	Certain

BLAST Alignment to Top Hit:

>CRA|18000004928524 /altid=gi|4503687 /def=ref|NP_001996.1| feline
sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian
sarcoma (PRCII) viral (v-fps) oncogene homolog; Oncogene
FES, feline sarcoma virus [Homo sapiens] /org=Homo
sapiens /taxon=9606 /dataset=nraa /length=822
Length = 822

Score = 1364 bits (3491), Expect = 0.0

Identities = 706/822 (85%), Positives = 716/822 (86%), Gaps = 70/822 (8%)

Frame = +3

Query: 72	MGFSSELCSPOGHGVLQQMQEAE LRLLEGMRKWM AQRVKS DREYAGLLHHMSLQDSGGQS	251
	MGFSSELCSPOGHGVLQQMQEAE LRLLEGMRKWM AQRVKS DREYAGLLHHMSLQDSGGQS	
Sbjct: 1	MGFSSELCSPOGHGVLQQMQEAE LRLLEGMRKWM AQRVKS DREYAGLLHHMSLQDSGGQS	60
Query: 252	RAISPDSPISQSWAEITTSQTEGLSRLLRQHAEDLNSGPLSKLSLLIRERQQLRKTYSEQW	431
	RAISPDSPISQSWAEITTSQTEGLSRLLRQHAEDLNSGPLSKLSLLIRERQQLRKTYSEQW	
Sbjct: 61	RAISPDSPISQSWAEITTSQTEGLSRLLRQHAEDLNSGPLSKLSLLIRERQQLRKTYSEQW	120
Query: 432	QQLQQELTKTHSQDIEKLKSQYRALARDSAQAKRKYQEASKDKDRDKAKDKYVRSLWKLF	611
	QQLQQELTKTHSQDIEKLKSQYRALARDSAQAKRKYQEASKDKDRDKAKDKYVRSLWKLF	
Sbjct: 121	QQLQQELTKTHSQDIEKLKSQYRALARDSAQAKRKYQEASKDKDRDKAKDKYVRSLWKLF	180
Query: 612	AHHNRYVLGVRAAQLHHQHHLQLLLPGLLRSLQDLHEEMACILKEILQEYLEISSLVQDE	791
	AHHNRYVLGVRAAQLHHQHHLQLLLPGLLRSLQDLHEEMACILKEILQEYLEISSLVQDE	
Sbjct: 181	AHHNRYVLGVRAAQLHHQHHLQLLLPGLLRSLQDLHEEMACILKEILQEYLEISSLVQDE	240
Query: 792	WAIHREMAAAAARIQPEAEYQGFLRQYGSAPDVPPCVTFDESLLEEGEPELPGELQLNE	971
	WAIHREMAAAAARIQPEAEYQGFLRQYGSAPDVPPCVTFDESLLEEGEPELPGELQLNE	
Sbjct: 241	WAIHREMAAAAARIQPEAEYQGFLRQYGSAPDVPPCVTFDESLLEEGEPELPGELQLNE	300

FIG. 2C

Query: 972 LTVESVQHTLTSVTDELAVATEMVFRRQEMVTQLQQELRNEEENTHPRERVQLLGKRQVL 1151
 LTVESVQHTLTSVTDELAVATEMVFRRQEMVTQLQQELRNEEENTHPRERVQLLGKRQVL
 Sbjct: 301 LTVESVQHTLTSVTDELAVATEMVFRRQEMVTQLQQELRNEEENTHPRERVQLLGKRQVL 360

Query: 1152 QEALQGLQVALCSQAK-----LQAQQE-----LLQTKLEHLGPGE----- 1256
 QEALQGLQVALCSQAK LQ + E LLQ E
 Sbjct: 361 QEALQGLQVALCSQAKLQAQQELLQTKLEHLGPGEPPPVLQLQDDRHSSTSSSEQEREGGR 420

Query: 1257 -----PPPVLQLQDDR-----HSTSS 1304
 PPP+ L+ + + HS
 Sbjct: 421 TPTLEILKSHISGIFRPKFSLPPLQLIPEVQKPLHEQLWYHGAIPRAEVAELLVHSGDF 480

Query: 1305 SEQEREGGRTPTLEILKSHISGIFR-PKFSNLYRLEGEFGPSIPLLDHLLSTQQPLTKK 1481
 +E +G + L +L + F NLYRLEGEFGPSIPLLDHLLSTQQPLTKK
 Sbjct: 481 LVRESQKQEVVLSVLWDGLPRHFIIQSLDNLRYRLEGEFGPSIPLLDHLLSTQQPLTKK 540

Query: 1482 SGVVLHRAVPKDKWLNHEDLVLGEQIGRGNFGEVFSGRRLRADNTLVAVKSCRETLPDDL 1661
 SGVVLHRAVPKDKWLNHEDLVLGEQIGRGNFGEVFSGRRLRADNTLVAVKSCRETLPDDL
 Sbjct: 541 SGVVLHRAVPKDKWLNHEDLVLGEQIGRGNFGEVFSGRRLRADNTLVAVKSCRETLPDDL 600

Query: 1662 KAKFLQEARILKQYSHPNIVRLIGVCTQKQPIYIVMELVQGGDFLTFLRTEGARLRVKTL 1841
 KAKFLQEARILKQYSHPNIVRLIGVCTQKQPIYIVMELVQGGDFLTFLRTEGARLRVKTL
 Sbjct: 601 KAKFLQEARILKQYSHPNIVRLIGVCTQKQPIYIVMELVQGGDFLTFLRTEGARLRVKTL 660

Query: 1842 LQMGDAAAGMEYLESKCCITHRLAARNCLVTEKNVLKISDFGMSREEADGVYAASGGLR 2021
 LQMGDAAAGMEYLESKCCITHRLAARNCLVTEKNVLKISDFGMSREEADGVYAASGGLR
 Sbjct: 661 LQMGDAAAGMEYLESKCCITHRLAARNCLVTEKNVLKISDFGMSREEADGVYAASGGLR 720

Query: 2022 QVPVKWTAPEALNYGRYSSESVDWFSFGILLWETFSLGASPPNLSNQQTREFVEKGGRLP 2201
 QVPVKWTAPEALNYGRYSSESVDWFSFGILLWETFSLGASPPNLSNQQTREFVEKGGRLP
 Sbjct: 721 QVPVKWTAPEALNYGRYSSESVDWFSFGILLWETFSLGASPPNLSNQQTREFVEKGGRLP 780

Query: 2202 CPELCPDAVFRLMEQCWAYEPGQRPSFSTIYQELQSIRKRHR 2327
 CPELCPDAVFRLMEQCWAYEPGQRPSFSTIYQELQSIRKRHR
 Sbjct: 781 CPELCPDAVFRLMEQCWAYEPGQRPSFSTIYQELQSIRKRHR 822 (SEQ ID NO:4)

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
CE00287	CE00287 PTK_Eph_orphan_receptor	285.3	7.8e-82	1
PF00069	Eukaryotic protein kinase domain	271.5	1.1e-77	1
CE00292	CE00292 PTK_membrane_span	240.7	2.1e-68	1
CE00290	CE00290 PTK_Trk_family	232.7	5.4e-66	1
CE00291	CE00291 PTK_fgf_receptor	190.3	3e-53	1
CE00031	CE00031 VEGFR	149.8	6.1e-45	2

FIG. 2D

PF00611	Fes/CIP4 homology domain	149.0	1.9e-42	1
CE00334	E00334 urotrophin_receptor	143.6	3.5e-45	1
CE00286	E00286 PTK_EGF_receptor	141.3	1.7e-38	1
CE00204	CE00204 FIBROBLAST_GROWTH_RECEPTOR	124.9	8.1e-34	2
CE00288	CE00288 PTK_Insulin_receptor	110.9	2.5e-29	1
CE00202	CE00202 EPHRIN_TYPE_A_RECEPTOR	108.3	1.9e-29	3
CE00203	CE00203 ERBB_RECEPTOR	81.8	2.1e-22	1
CE00549	CE00549 NGFR	69.2	7.4e-20	1
CE00201	CE00201 EPHRIN_TYPE_B_RECEPTOR	57.0	5.9e-15	4
CE00359	E00359 bone_morphogenetic_protein_receptor	26.3	9.4e-07	1
PF00017	Src homology domain 2	10.4	0.024	1
PF00422	ATP synthase Alpha chain, C terminal	7.7	0.11	1
CE00289	CE00289 PTK_PDGF_receptor	-33.8	8.2e-05	1
CE00016	CE00016 GSK_glycogen_synthase_kinase	-290.0	1.2	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00611	1/1	1	94	1	109	149.0	1.9e-42
PF00422	1/1	210	234	111	135	7.7	0.11
PF00017	1/1	441	460	60	79	10.4	0.024
CE00201	1/4	497	510	662	675	0.1	17
CE00202	1/3	497	511	680	694	3.2	2.4
CE00202	2/3	534	557	721	744	5.3	0.63
CE00201	2/4	534	567	702	735	4.2	1.3
CE00031	1/2	483	580	868	972	4.5	0.17
CE00289	1/1	489	588	1	109	-33.8	8.2e-05
CE00204	1/2	600	636	649	685	68.1	3.5e-18
CE00201	3/4	598	636	766	804	21.3	2.9e-05
CE00359	1/1	495	636	145	299	26.3	9.4e-07
CE00204	2/2	653	692	702	741	56.1	7.1e-15
CE00201	4/4	653	697	824	868	30.5	9.5e-08
CE00334	1/1	490	734	539	803	143.6	3.5e-45
CE00203	1/1	598	737	850	989	81.8	2.1e-22
CE00202	3/3	598	740	827	970	99.5	5.2e-27
CE00290	1/1	491	744	1	282	232.7	5.4e-66
CE00291	1/1	491	744	1	285	190.3	3e-53
CE00286	1/1	491	744	1	263	141.3	1.7e-38
CE00292	1/1	491	744	1	288	240.7	2.1e-68
CE00031	2/2	598	744	1056	1203	145.3	1.4e-43
CE00287	1/1	491	744	1	260	285.3	7.8e-82
CE00288	1/1	491	744	1	269	110.9	2.5e-29
CE00549	1/1	598	745	693	840	69.2	7.4e-20
PF00069	1/1	491	746	1	278	271.5	1.1e-77
CE00016	1/1	431	751	1	433	-290.0	1.2

FIG. 2E

```

1 CTGGCCACCA GGCTGGCGCA GCCAAGGCCG AAGCTCTGGC TGAACCCTGT
51 GCTGGTGTCC TGACCACCCT CCCCTCTCTT GCACCCGCCT CTCCCGTCAG
101 GGCCCAAGTC CCTGTTTTCT GAGCCCGGGC TGCCTGGGCT GTTGGCACTC
151 ACAGACCTGG AGCCCTGGG TGGGTGGTGG GGAGGGGCGC TGGCCAGCC
201 GGCCTCTCTG GCCTCCACC CGATGCTGCT TTCCCCTGTG GGGATCTCAG
251 GGGCTGTTTG AGGATATATT TTCACTTTGT GATTATTTCA CTTTAGATGC
301 TGATGATTTG TTTTGTATT TTTAATGGGG GTAGCAGCTG GACTACCCAC
351 GTTCTCACAC CCACCGTCCG CCCTGCTCCT CCCTGGCTGC CCTGGCCCTG
401 AGGTGTGGGG GCTGCAGCAT GTTGCTGAGG AGTGAGGAAT AGTTGAGCCC
451 CAAGTCCTGA AGAGGCGGGC CAGCCAGGCG GGCTCAAGGA AAGGGGGTCC
501 CAGTGGGAGG GGCAGGCTGA CATCTGTGTT TCAAGTGGGG CTCGCCATGC
551 CGGGGGTTCA TAGGTCACTG GCTCTCCAAG TGCCAGAGGT GGGCAGGTGG
601 TGGCACTGAG CCCCCCAAC ACTGTGCCCT GGTGGAGAAA GCACTGACCT
651 GTCATGCCCC CCTCAAACCT CCTCTTCTGA CGTGCCTTTT GCACCCCTCC
701 CATTAGGACA ATCAGTCCCC TCCCATCTGG GAGTCCCCTT TTCTTTTCTA
751 CCCTAGCCAT TCCTGGTACC CAGCCATCTG CCCAGGGGTG CCCCTCCTC
801 TCCCATCCCC CTGCCCTCGT GGCCAGCCCG GCTGGTTTTG TAAGATACTG
851 GGTTGGTGCA CAGTGATTTT TTTCTTGTA TTTAAACAGG CCCAGCATTG
901 CTGGTTCTAT TTAATGGACA TGAGATAATG TTAGAGGTTT TAAAGTGATT
951 AAACGTGCAG ACTATGCAAA CCAGGCCAG TCTCCAGTGT GGTACCGTTG
1001 CTCCTGCATC GCAGCTGAGG ATAGGGGGCC AGTTAGGCCT ACACAGTGGC
1051 CTGCCTGCCT GGATGTGGGC CCAAGTCAGA AGGCCAAAGT CCTCCAAGGG
1101 GCGGGAGGAT GCGCCAGCCC CTAGTGGAGG AGCTGGTGCC CCTGGGGTGG
1151 GGCTGGTGAC CCCTGGTCCT CAGGAGCTGA GCACTAAACT CCCAAAGTCC
1201 TGGTTTCCAG CAGTGTGAAG AACTGGGCCT ATTGTGTCTT CCTGGGCTGA
1251 AGTGATCTGG TCGCCACAGG CTATAGGGCT GAGGCCTAAG GTGGAGGGAG
1301 GCCTGACTGA ATCAAGATGA CTTCTTGTGG GGAGCCTGAG TCCCAAATGG
1351 AAAACTCCAC GCCTGTCCGC TCCCCAACCC CTGCCCCTTG ATTTCCCCAG
1401 GTCTCCCTTG GGACAGGAAG CCCCTGCCTG GGGGTAGGAG GATGGGGACA
1451 AAACCACTAG GATCTGTATC CGAGAAGCAG TCTCTGTTCG GGATATTTAC
1501 TTGGAATTT TATTCAAATG GAAGCTGGCG CCTGAGCCTC TCCTTAGGGA
1551 ATTCCTGAG GTGGGGAGGG CTGGGACCAG GGTTCCCTCT TTCTTCTG
1601 CGGTGGCCCT GGCCTGGTGC TAGGACTGCG CGCCTCCCCT CAGTACCCGC
1651 GGACACCCTG GGCTTCCCTG GGCCAGCAT CTGCCTGGGG CCTCGCCCTG
1701 GGCTCCCCCT CCTGACCCCC ACCTTGCGCC CTTCCCGGT GTTCCCGGGG
1751 CGCTGCCGGG CCCTGGGGCC TGCGGGGCGC GGGCGGCTCT TGGCTGGGCC
1801 ATTCTTTCCC GGCCCCCTCC TCCCTTCCGT TTCCGTGGCC GTGCGGCCGG
1851 CTAGAGGCTG CGGCCAGCG CGGAGCAGGG GGGCTGGCAG GCGTCGGGAC
1901 GGTCGGGCCG GTCCCGCCCG CCCCTTCCCC TCCACAGGCC CGCCCCGGGG
1951 CCTGGGCAA CTGAAACCGC GGGAGGAGGA AGCGCGGAAT CAGGAACTGG
2001 CCGGGGTCCG CACCGGCCT GAGTCGGTCC GAGGCCGTCC CAGGAGCAGC
2051 TGCCCGTGCG GGTACCTCTA GCCCCGGGGC CTGGAGGAGC GGTGGGAGCT
2101 GGGGGCGCGG CAGGCAGGGG CAGAGCAGGC GTTCCGAGGG CCAGAGACCC
2151 ACCCAGGTGG GGGTAGGGGC CGCGGAAGGG CGGGGATGGC CGCAGGGGCA
2201 GGGCTCAGGC TGTGGGCGCC TGAGGCTTCA GCTGGGGCAG GCTTGGCCTG
2251 TCGAGGACCT GGGCAAGGGT GTCCCTGTAA GGGGTGGTGG GTGGAAGGGC

```

FIG. 3A

2301 CTGGGGAGGG AGGCTCCAGG TTGGCTCCTG TTCCCGAACG TGCGGAGGAG
2351 ACCCTGACGC TAAGGAAGCA ATGAGGGCCA GTCCCCAGGC CAGGCTGCTG
2401 CTGGGTACCC ATGGCTGCGT GTGAGCGAGG CAGGACCCCA CCTCCTCCCC
2451 GTCTGCAGTC CATCCTGACC CTACAGTCCC CAGCCTCCTC GTCCCATGCC
2501 TCCGTCTCCA GCTGCTGCCT TGCCTCCAGG GATGGCCCCT TTTCTGTCCC
2551 CAGAACAGCA CTATGGGCTT CTCTTCTGAG CTGTGCAGCC CCCAGGGCCA
2601 CGGGGTCTCTG CAGCAAATGC AGGAGGCCGA GCTTCGTCTA CTGGAGGGCA
2651 TGAGAAAGTG GATGGCCCAG CGGGTCAAGA GTGACAGGGA GTATGCAGGA
2701 CTGCTTCACC ACATGTCCCT GCAGGACAGT GGGGGCCAGA GCCGGGCCAT
2751 CAGCCCTGAC AGCCCCATCA GTCAGGTGGG TCTCTATGGG ACTCTGGTGG
2801 GTGCTGGCGT ATCTGCCTTC TCCTTCCTCT CCTGGGGGCC CTCTGGGGCA
2851 GTGGCTGGAG ATCTGGCAGG CCAATGCTTG GGAGCCATTG TGCCCCCCTC
2901 CCTGCCTCCC CCATCTGTGC TGTATAGTCC TGGGCTGAGA TCACCAGCCA
2951 AACTGAGGGC CTGAGCCGCT TGCTGCGGCA GCACGCAGAG GATCTGAACT
3001 CAGGGCCCCCT GAGCAAGCTG AGCCTGCTCA TCCGGGAACG GCAGCAGCTT
3051 CGCAAGACCT ACAGCGAGCA GTGGCAGCAG CTGCAGCAGG AGCTCACCAA
3101 GGTGAGCGGG CAGCACTGGG GCTTCGGTCA TTTCTGTCTA AATTTTGAGC
3151 CTCGAAGGGG TTGTTTTGCA CAAGAGGCCC TGGATTCACT GGGGAAGTGT
3201 AAGTCCCTGA CCGCAGGCCT GGCTTGCTCT AACCTTGATG TAGCTTCCTC
3251 TCTTCCTTCC CCTACGTTGA GCTGGCTTGC AGCAAGGCCT CTCTGTGCTT
3301 TTTCTGTGCC TGGGCAAAGT GCTGGGAGTG TAAGGATGAG TGACCGGTCA
3351 CGTGCCTGGG AGAAGCTCAG AATCGGTACT CGCCTCCACA CTGTGCCATC
3401 TGGCTCTGGG TTCTGAGAGT CAGGGAGAGG AATGAGGGTC AGTCTGTTTG
3451 CCTTCGACCT ATGCAGCCTC CTCTCAGGGC CCCAGAGACT GGGCAGCAGC
3501 ATGGCCCCCC GAAGGTCGAG GACTCGGGCC GTGAAGTCAG CCTGCCTAGG
3551 TTTGAATCCC ACCCAGCTCC TCAGTCTAGA GGCTGTGTGA TTTGGAAC TA
3601 TTTATCTGGG AGCCTAGTGC CCCCATTAGG TGTGCTGGTC ACCCTCCCTG
3651 CACCACACCC CTTCTCAAG TGCAGAGCCC AGCCTTGCCA TGGACCCACA
3701 GCGGCCCTG GTGGCCACCC TGGCCCCATT CTCGCCCCA AAAGATCATC
3751 TGATTCAAGG GTGGGCCCAT TTTTATAAAG TTTTGCTGGA ACACAGCTAT
3801 GCCCCTTTGT TTTTATATTG TCTGTGACTA CAATGACAGA GTTGAGTAAT
3851 TGTGACAGAG GCTCTATGGC CTACAAGCCT AAAATATTTA TTTACTATCT
3901 GGCCCTTTAA GAAAAAGACT GATCTAGTCG AGGAATCTAG CTCAGTTACA
3951 GATGGGGAAA CTGAGGTTGG GCGCTTGCCC AACATATCCC AGCACATAAA
4001 CAGGAGAACT GGGACGAGAA CACTGATCTC GGGCTGTCAT CTATTCCTAC
4051 TGCCAAGAAC ATAATTTGCA GGACCCAGTG CAAAGTGAAA TTGTGGGGGT
4101 CTTTGTTAAA AGATTGCTAG GAATTTCCAG GTGGCAATAA TGGAGAATGA
4151 AACCAAGCAC AGGGCCCTTC TACATGTGGA GCGGCTGTG ACTGCACAGG
4201 CCGTGACAC CTGCAACTGG CCCTGCCTGC CACCAGGCTA CCACTGTCAG
4251 TCCAAGGAGG GACCGTTGTA GCCTGTAGTC TACCTCTTTG CCTCCCCAAG
4301 GGGTCTGTCT TCAACAGGCT CTCTGATCTT TGA CTCTCAC GTCAGCAGCC
4351 AGCTTTCCCA GAAGTCTCCA GGTGCTCCTT GCCTGACGAC AGGACCTTTC
4401 CAGGGCTTCA CCCCAGGCAA GAATCTTCCA CAACTGGGGA CCTGCTGCCC
4451 CAACTGGGCC TCTCCTCTCT CCCTAGACCC ACAGCCAGGA CATTGAGAAG
4501 CTGAAGAGCC AGTACCGAGC TCTGGCACGG GACAGTGCCC AAGCCAAGCG
4551 CAAGTACCAG GAGGCCAGCA AAGGTTCTGT GCTTCCCTTG CTGGCAGGGA

FIG. 3B

4601 GGGAAATCCGA AGCCAGTGCT GACCTGTCCT TGGGTACCCA GAGAGTGGGG
4651 GCTGCCTGGG CCTCCATGCT GTCATCTATA CCCCTTGCCC CCCTTCTGGC
4701 AGACAAGGAC CGTGACAAGG CCAAGGACAA GTATGTGCGC AGCCTGTGGA
4751 AGCTCTTTGC TCACCACAAC CGCTATGTGC TGGGCGTGCG GGCTGCGCAG
4801 CTACACCACC AGCACCACCA CCAGCTCCTG CTGCCCCGCC TGCTGCGGTC
4851 ACTGCAGGAC CTGCACGAGG AGATGGCTTG CATCCTGTAA GCCCCGAGCC
4901 CCGTCCCCTG GCCCCACCC TTGAGCAGCC CTAAGCCCAG CCATCAGGCC
4951 CAGAGGCAGG ACCCAGAAAA TCCATTGCTG GGAAGGTGCT GGCCATGTAA
5001 CCACATGAGA ACGGGACCTG GGCCAAGGAT TGGAAACAGG CAACTTACCT
5051 CTGAATTACA CTATTCCAGG GTCTCATTAT TCCAGGGTTT TATTACATTC
5101 ATTGAGCACT GTTCTGGGCT CTGGATTATA CCAGAGAACG ATGGTAGACA
5151 AAAACATCTG TCCTCAGGGA TCTTTCGTGT TAGTGGAGTG AGAATGTGAG
5201 GAGCACTAAG AGCCATGGAG AAAAATAAAG CAAGAGAAGT GGATCGGGAC
5251 CTGGGAGCAC GGAGGCAAGG GAGGAGGTGA CAGTTGTCCA TAGAGTGATC
5301 TGGGAAAGCC TCTTGAGAGG TGACATTCAA AGAGGCCCTT GAGAGGGGTA
5351 CGGGAGTGAA TCATGGGGCT ATTTGGAGAA AGACCATTCC AGAAAGGAGG
5401 ACAGCAATTA CACAGGCCCT GAGGTAGGAG AGTACCAGGG ACTAATAGCC
5451 AGGAACCAGT GGTGCCTCTG AGAGTGAGGG AGGGGGAGAG TCATACACGA
5501 GGCTGGAGGA GGCAGGCGTC AAGGGCTACT GGGTGATAGA AGGTCTAGCA
5551 GGGCCATGGT GAGGACTTTG GCTCTGGGTG AACAAGAATG GCATGATCTG
5601 ACCTCTGTTT TTTTGTTCCT TTTTGTTCCT ACTTTTTTTG AGTCAGAGTC
5651 TCGCTCTGCC GCCCAGGCTA GAGTGCAGTG GCATGATCTC GGCTTACTGC
5701 AACCTCCGCC TCCCAGGTTT AAGTGATTCC CCTGCCTCAG CCTCCCGAGT
5751 AGCTGAAACT ACGGGCATGC GCCACCACAC CCAGCTAATT TTTGTATTTT
5801 TAGTAGAGAC GGGGTTTCAC CATGTTGCCC AGGCTAGTCT CTAATTCCTG
5851 GGCTCAAAGC GATTTGCCTG CCTCTGCCTC CCAAAGTGCC GGGATTACAG
5901 GCATGAGCCA CCATGCCCAG CCCTGACCTC TGTTTTAATA AGGCCACTCT
5951 GGCTGCTGTG CTGCAAATAG ACTTCAGGGA GCAAGGACAG AAGCTGGGAG
6001 GCCAGAGAGC AGGCTGCTTG CCATAATCCA GATCCAAGCT TTTGGCCAGC
6051 TAGGACGGGG AGGTAGCAAT GGAGGTGAGG CGCGGTCAGG TCCTGGGGCA
6101 GGTCTTGAA GGTGAAGCCA GTGGGATTTT CCTATGGATT GGAAGTGGGG
6151 CGTGAAATAG AGGAGTCAGG GGTCACTCTG GGGATTTGGC CTGGAGCAGC
6201 TGGAAGATGG AGTGCTGTT AACTTATGTA GGAAGGCTG TGGGAAGAAG
6251 AGGTTTAGGA GACAAGGATA GCAGTTCATT TATTTATTTA TTTATTTATT
6301 TATTTATTTA TTTATTTAGA GATGTAGTCT CATTCTTTCG CCAGGCTGGA
6351 GTGCAGTGGC GCGATCTTGG CTCACTGCAA CCTCCACCTC CCAGGCTCAA
6401 GCGATTCTCT TGCTCAGCC TCCCGAGTAG CCAAGTAGCT GGGACTACAG
6451 GCATGTGCCA CCATGCCTGG CTAATTTTTG TATTTGCTTT TTCAGTAGAG
6501 ATGGGGTTTC ACCACGTTAG CCAGGCTGGT CTCGAAGTGA CCTCAGGCAA
6551 TCCACCCGCC TCGACCTCCC AGTGTTGGTA TTATAGGCGT GAGCCACTGT
6601 GCCTGGCCCA CTGGATCCTT ATTACAACCTG CCAGTGTCCC TCTTATATAT
6651 ATCAGGAAAT AGAAGATTAG GGAGAGGTTA AATAATTTGC CTAGAGTGGC
6701 ATGGCTAGCT CGAAGTGAGG CAGGGGTCAA CCCAGCCCT GACTCCAAAC
6751 CCAGGGTCTT AGGCCTGAAC TGCCCAGCCT TGCCCAGCCT GAGGCTCCCC
6801 TGAAGTGGGA TCCCGTCTCG GGGGAGGAA GGAGATCCTG CAGGAATACC
6851 TGGAGATTAG CAGCCTGGTG CAGGATGAGG TGGTGGCCAT TCACCGGGAG

FIG. 3C

6901 ATGGCTGCAG CTGCTGCCCG CATCCAGCCT GAGGCTGAGT ACCAAGGCTT
6951 CCTGCGACAG TATGGGTAAG CCCCCTCCTT GCTCCTGCTG GGCCAGGGC
7001 TGCTGGCCTG TCCACTGACG GGGCGCTGTC CCCCACAGGT CCGCACCTGA
7051 CGTCCCACCC TGTGTCACGT TCGATGAGTC ACTGCTTGAG GAGGGTGAAC
7101 CGCTGGAGCC TGGGGAGCTC CAGCTGAACG AGCTGACTGT GGAGAGCGTG
7151 CAGCACACGT GGGTGGTGGC TTTGCACCTG GGCTGCGGCG GGGCTCCCAG
7201 CAGACCACGA GTGTTTATGT AGGCAGGGCT AGGTCGTGGA GACTGTCCAC
7251 ACAGAGCTGT CACCAGGTGG CCGGGCTTGC TTGGCTCTAC AGGGATGCAC
7301 TGGACCTGGG TTGAGGGGGC AGGAGGGCTC GGTTCTAATG CTGCCCTTCT
7351 CTTGGGTGCA GGCTGACCTC AGTGACAGAT GAGCTGGCTG TGGCCACCGA
7401 GATGGTGTTT AGGCGGCAGG AGATGGTTAC GCAGCTGCAA CAGGAGCTCC
7451 GGAATGAAGA GGAGAACACC CACCCCCGGG AGCGGTGAGT GGGCCCCTGC
7501 CTGCAGCAGC CTCCTGGGCC TCCCTCCCTC CTACCTACCC TAACTGCTGC
7551 TGGCTAGCCG CCGCAGACCG AGCCCTTATT CTTTCATCCAC CCTCCCACCC
7601 GCCCCTGCCT GCAGGGTGCA GCTGCTGGGC AAGAGGCAAG TGCTGCAAGA
7651 AGCACTGCAG GGGCTGCAGG TAGCGCTGTG CAGCCAGGCC AAGCTGCAGG
7701 CCCAGCAGGA GTTGCTGCAG ACCAAGCTGG AGCACCTGGG CCCCAGGCGAG
7751 CCCCCGCTG TGCTGCTCCT GCAGGATGAC CGCCACTCCA CGTCGTCCTC
7801 GGTGAGCTGC CCCATCCGCG GCCGCTGCCC GCCACCGGCC TGCCACCTG
7851 GGGCTGCGCT CCTCATTTTC GCCCTCCCCC TCCCTAAGCC TGGCCACCCG
7901 CTGACGTCTG TCCCTGGCCT CAGGAGCAGG AGCGAGAGGG GGGAAAGGACA
7951 CCCACGCTGG AGATCCTTAA GAGCCACATC TCAGGAATCT TCCGCCCCAA
8001 GTTCTCGGTG AGTGGCGCCC AGCCTGGGCC CCCCTACTGT TGTGTTTCGA
8051 GTTTAATCAC TGGGATGTCC TAGAGAGGAG GCTCTGCCCA GGCTGCTTGT
8101 ATTGGGAAGT TCCTCTCTTC CCTGGGATTC CAGGCTGCAG ATGTCCCCAG
8151 ACCCTGCCCC TGTGACCCCT CCCTTTCCAT CGCCCCAGTG TGCTAAAGGG
8201 ACCAGCAACC TCGACTATTC CATGGCTCTC CCTGCTTCAG GAGCGGTTGG
8251 GGGCCTGTGG CCTGGAGGAG GAGGCACCAG CTTGGTTTGG GGTCTTCCTG
8301 CCTGGGCTTC CCTTCCCAGC TCTGCCCAGC GTGAGCCTGG GCCAGTCCAG
8351 TGCCCACTCC AGGGGCCTGT GGATGGCTCT GCATGCCACT CCATGGTTGT
8401 AAGGGCTGAG GGCATATAGG GGGGAGAGAG AGACCCCCGG CTGCCCCCAC
8451 GGCCTCTTCA ACAAGGTGGT TAAGTGACTC CTCCTCGATC CTCCCTTGCC
8501 CAGCTCCCTC CACCGCTGCA GCTCATTCCG GAGGTGCAGA AGCCCCTGCA
8551 TGAGCAGCTG TGGTACCACG GGGCCATCCC GAGGGCAGAG GTGGCTGAGC
8601 TGCTGGTGCA CTCTGGGGAC TTCCTGGTGC GGGAGAGCCA GGGCAAGCAG
8651 GAGTACGTGC TGTCGGTGCT GTGGGATGGT CTGCCCCGGC ACTTCATCAT
8701 CCAGTCCTTG GATGTGAGTG GGGCTGGGAC CCGAGCCTTC CAGGCCTCAC
8751 TCTTCCCCTC CCTTCCCTTC CCAAGGGAA ATGGCCTTTC AGGGTAGGGG
8801 GTAGCTGCCA GGTCTTGGAT GCCTCCCTAG CAGGGCTGGC TGGAAGGGGC
8851 CACAGAGACC ACCCTGTCCC TGCAACAAA TAGAGGCTTA AGTGTGAGTC
8901 CTCCCCTGGT GGGGCAGCAG GATGTCATGT GCCATCAGAT GGCATCTTTT
8951 CTGGAGGTCT CTCTGCCCCT GGTCTGGGC AGGCCCTTTC TCCCCTGCTG
9001 CTCTCCCTTT CCCCCTCCCA GGGCTCACGC CCCCTCAGAA TGGAGGCTGC
9051 TGACCCCCGGG TCCCCTGCCC TGCAGAACCT GTACCGACTG GAAGGGGAAG
9101 GCTTTCCTAG CATTCCCTTG CTCATCGACC ACCTACTGAG CACCCAGCAG
9151 CCCCTCACCA AGAAGAGTGG TGTGTCTCTG CACAGGGCTG TGCCAAGGT

FIG. 3D

9201 GAGCCTGCAC CCAGCCTGGC CCATGCCACC TGTGGCAGGG CTTGGGGAGT
9251 GTGGGTCAGG CCCACCCAGC GTCTGAGCAG AAAGGGCTTT CCAGGCCCTC
9301 CGTCTACATA CAAGATGCAG AGTGAGTGAC CCTCAGGGCC AGCCTTGCTC
9351 TAGGTTTGGG ATGTCAGGGC CACTCCTATG CCATGGGCTG TACACACCAG
9401 GTTGGTGCTT ACCTGGTCAG GGCACCTGCC TGGACCCCGT AGTCATCTCA
9451 GTGTGCTCCC CACGTGGTCC CACCCCTGGT CACATATGGA GGCGCCAAAA
9501 AATGGAGGAC ACAGCCCTTC TAAGGGCCCA GCACCCCTTT TCTTCAGACT
9551 TCTGATCCCC TGTCTCCTCT CTTCCCCAGG ACAAGTGGGT GCTGAACCAT
9601 GAGGACCTGG TGTGGGTGA GCAGATTGGA CGGGTGAGTG CGCCTCTGCT
9651 GGCCTCCTTG TCGCTGGCGA CTTCTCCTGA GTCGCGCCTG GGCCCCCTGC
9701 CCTACCACCC AGAAACCTCC CTGCCCCATC TGATTCCCCA CTTGTACCCC
9751 GACTCCCTGC CCAGCCCCCA CCACACACCA TCCTCCAGGA AACGGGACAG
9801 TACCTACGCT GAAAACTCCC AGCAGACAGC TCTGCCAGCA CCCTGACCTC
9851 ATCACCCCA CCCAGGCCGC CCCCATCGAG CTCTTGTGTG CACGCAGGGA
9901 GACACCCTGT TACTGTAAGC CATAAGATAC CTGTTTAGGG AAGAAGTCAC
9951 TGTCTAAAA ATCAGAATGC TTTTCAAACC CAAGGGAGAG TGATTTTTTG
10001 ATTTCCATGT CACTTCTCTC AGGAAGGGTG GCACATCGGA GGCAACTTTC
10051 CCTGCCTGCC CCATGTGCTC TCTAGGTTCC CCAGCGAGGG TCAAACCTCC
10101 AGAGAGCCTG GGTGGAGGGG TCCGAACACG GGGGCCCTC ACCCAGGGGT
10151 AGGAAGCAGA ATGGGTAGGA AGCGGAGAAG AGAACTGCGG GACTGGGAAG
10201 GCCGTGGTAG GAGCCCAAGA CCGTTTCAGG GGAACCTTGG CGAAGTGTTT
10251 AGCGGACGCC TGCAGCCGA CAACACCCTG GTGGCGGTGA AGTCTTGTCG
10301 AGAGACGCTC CCACCTGACC TCAAGGCCAA GTTTCTACAG GAAGCGAGGT
10351 GGGTGATAAA CTAATGATCA CCACGGGTCC CGCATACACA GAGGTTACAC
10401 TGCATGGCAC AGTGTGAAGT GCTTGACCAC CGTGGTGGTG TTTAGTCTC
10451 GAGGCCCCC ATTGCGGGTA GTACCCCTT ATAGTGCCGA AGGGTAGAGG
10501 CTGCCCCAGG TCACACGTCC GGGTCTGCTG GCCTTGAGG CCAAGCTCTT
10551 CTCCCATCAT CCCTGGGGGG CCCTGGGGAG GCGGGCCTGG CCACGTAGAT
10601 CCTGAGCAGC AGTGCCCTCC AGGATCCTGA AGCAGTACAG CCACCCAAC
10651 ATCGTGCGTC TCATTGGTGT CTGCACCCAG AAGCAGCCCA TCTACATCGT
10701 CATGGAGCTT GTGCAGGGTG AGCGCGGGG GCTGAGCTCC AGGTAGGGCG
10751 CGCAGCCTGG TCAGGTGGCA GCCTTACCTC AGGAGGCTCA GCAGGGGTCC
10801 TCCCCACCTG CAGGGGGCGA CTTCTGACC TTCCTCCGA CGGAGGGGGC
10851 CCGCCTGCGG GTGAAGACTC TGCTGCAGAT GGTGGGGGAT GCAGCTGCTG
10901 GCATGGAGTA CCTGGAGAGC AAGTGCTGCA TCCACCGGTG AGTGGGCGGT
10951 GGCCACGGGC CCTGCCAACA CCCCCACCA GAGTCAAGAG GTACCTATAC
11001 CCCTAGGGCC CCCCCTGGA CCATCAGGCA TCAGCTCCAG AGGGGGAGTT
11051 GGCCTCTGTG GTAGACAGGG GTGCCAGGG CCGGGAGCAG CTTTTGTCTT
11101 TGGCTTTTCT AGAGTGTTCA GCCAGGGCTG GGCAGGCGAC TGTTGGCCAA
11151 ATGAGCCCCT GCCCTGTCTC ACCCAGGGAC CTGGCTGCTC GGAAGTGCCT
11201 GGTGACAGAG AAGAATGTCC TGAAGATCAG TGACTTTGGG ATGTCCCGAG
11251 AGGAAGCCGA TGGGGTCTAT GCAGCCTCAG GGGGCCTCAG ACAAGTCCCC
11301 GTGAAGTGGA CCGCACCTGA GGCCCTTAAC TACGGTACCT AGTCCCTGTC
11351 TACCCTGGAC TCCATGGCCA GAGGCCAGGC CTGGGTCCTG CCGGCTGCCT
11401 CGCCCTGGCC CCAGGGAGGG TGCATCACG CTGCCTCACC TCCTCGCCTC
11451 CTCTGCAGGC CGCTACTCCT CCGAAAGCGA CGTGTGGAGC TTTGGCATCT

FIG. 3E

```

11501 TGCTCTGGGA GACCTTCAGC CTGGGGGCCT CCCCCTATCC CAACCTCAGC
11551 AATCAGCAGA CACGGGAGTT TGTGGAGAAG GGTAAGCACC CTGTGATGAC
11601 AGCAGCCTCA GGCTGCACCC TCTTCCAGAT GCTCCAGCCG GACTCTTCTA
11651 ACTCCCTTAA TGCCAACCTT CCCACCAGGC AGAATAAGAA TAACCTGGCC
11701 AGTTGCTCAC GCCTGTCATC CCAGCACTTT GGGAGGCTGA GCTGGGTGGA
11751 TCACTTGAGC CCAGGAGTTC AAGATCAGCT TGGACAACAC AGTGAAACTC
11801 CATCTGTACA AAAAATACAA AAATAGACTG GGCACGGTGG CTCACACCTG
11851 TAATCCCAGC ACTTTGGGAG GCCGAGGCAG GTGGATCACC TGTGGTCAGG
11901 AGTTTGAGAC CAGCCAGACC AACATGGTGA AACCCCATCT CTAATAAAAA
11951 TACAAAAATT AGCCAGGCAT GGTGGCACGT GCCTGTAATC CCAGCTACTT
12001 GGGAGGCTGA GGTGGGAGAA TTGCTTGAAC CCAGGAGGCG GAGGCTGCAG
12051 TGAGCCGAGA TTGTGCCACT GCACTCCAGC CTGGGCGACA AGAGTGAAAC
12101 TCCATCTCAA AAAAAACCAA AAAACAAAAA ATACAAAAAT TAGCTGGGTG
12151 TGGTGACATG CGCCTGTAGT CCCTGCTACT CGGGAGGCTG AGGTGGGAGG
12201 ATCACTGGAG CCCGGGAGGT GGAGGTTGCA GTGAGCTGAG ATCATGCCAC
12251 TGCACCCCAA CCTGGGTGAC AGAGAGAGAG AGAGACCTTG ACTCGAAAAA
12301 GAAAAAAACC TGGGCGCAGT GGCTCACGCC TGTAAATTTCA ACATTTTGGG
12351 AGGTGAGGA AGGTGGATCA CTTGAGTCTA GGAGTTTGAC ACTAGCCTGG
12401 CCAACATGGC AAAACCTGTC TCTACTAAAA ATACAAAAAA TTAGCGAGGT
12451 GTAGTGGTGC AAGCCTGTAA TCCCAGCTAC TTGGGAGGCT GAGGCACAAG
12501 AATCGCTTGA ACCTGGGAGG TGGAGGTTGC AGTGAGCTGA GATCACACCA
12551 CTGCATTCCA GCGTGGGTGA CAGAGCAAGA CTCCATCTCA GAAAAAGAAA
12601 AAAAAAATA GAATATCCCT GTAGCTACTA CTGAGTGAGC ACCTGGTCTG
12651 TGCTAGGTCA CATGTTATTT CATTTGCTCA TCACTACATG TGTGGTAGGG
12701 ATTAATATGT CCCTTTCTCA GATGGAAAAA CAGGCTGGCA GAGGGGACAC
12751 AGCTAGCACG TGGTAGGATT AGGATCAGAA GCCAGGCCTC TTTGTCCTTT
12801 GGGCCCTTGG TGGAGAACAG TGCATCCTTC AGAACAGTGC ATCTTAAGCA
12851 GCTCCTATGG CTCATGGTAT CCCCCAGAGT CTGCCGAGGA CCCTCAAACCT
12901 CCCTCCTCAT GCCTGGTGTG CTGTGCCTCT CCTCACAGGG GGCCGTCTGC
12951 CCTGCCCAGA GCTGTGTCCT GATGCCGTGT TCAGGCTCAT GGAGCAGTGC
13001 TGGGCCTATG AGCCTGGGCA GCGGCCCAGC TTCAGCACCA TCTACCAGGA
13051 GCTGCAGAGC ATCCGAAAGC GGCATCGGTG AGGCTGGGAC CCCCTTCTCA
13101 AGCTGGTGGC CTCTGCAGGC CTAGGTGCAG CTCCTCAGCG GCTCCAGCTC
13151 ATATGCTGAC AGCTCTTCAC AGTCCTGGAC TCCTGCCACC AGCATCCACA
13201 CTGCCGGCAG GATGCAGCGC CGTGTCCCTCT CTGTGTCCCT GCTGCTGCCA
13251 GGGCTTCCTC TTCCGGGCAG AAACAATAAA ACCACTTGTG CCCACTGAAC
13301 ACTCCTGGCA TGTGCACTCC TCTGGAAGGC AGGTCTCAGA AGGCACAAGT
13351 GCCGGTATGG TGGCCTTGGG GAAGGAGGAG GACAGGCAGT ATGCATGGGG
13401 CAGAGCTGAC ATGATTTAGT AGCAGCTGGA TGTGAGACAT GCGGAAGGCG
13451 GGGGAGAGAT CAGGATGATA TACAGGCTAT GGCCAGATGG CGGTGTCATC
13501 CCCTGAAATA GGATTATAGG AAGAGGATCA GAGCTTCGAG GAGGATGTTG
13551 AGTTTAGAGA TGTTGCATTT TATTGGAGAT AAAAGTGTGG GTGAAGCCAG
13601 GTGTGGTGGT AGACACCTGT AGTCCCAGGT ACTTGGGAGG CCAAGGCATG
13651 TGGATTGCTT GAGCCTAGTT TGAGACCAGC CTGGGCAACA TGGCAAAACT
13701 CCATCTTTAC AAAAAACAAA AACAAAAAAC AAAAAACCAA GTAAATTAG
13751 CCAGGCGTGG TGGCACACAC CTATAGTCCC AGCTACTCAG AAGGCTGAGG

```

FIG. 3F

13801 TAGGAGGATC AATTGAGCCT CGGAGGTCGA GGCTGCAGTG AGCTGTGATC
13851 ACACCACTGC ATTCCAGCCT GGGCAACAAA GCGAGGCCCT GTCTCAAAAA
13901 TAAGTAAATA AAAATAATAA ATAATTAATT TAAAATGTAG ATGAATAGGT
13951 CTGGAAGCCC AGATGGAGAT GAAGGCTGGC AATAGATGTG TGAATCATTG
14001 GCTTATGAAT ATTAGAGAGT AGCTGACACT ATGGATGCGT ATAACACTCG
14051 CATAAAATTC AGGAGGAGAT GAGAAGAGAG TTCCACTCAA AGAAGACTGA
14101 TGTGGCTGAT GAGGAAGAAA ATGCTTTTGA GGGAGTTGTT TCTCAAGATG
14151 AATTTATTGA GGAATAAGAT GGCAGACTGG GGAGCCTTCA CCTCCTCCCC
14201 TAAGTCCAG TGAAACCTAA AAAGTCATCT GAAATATTAA CATCACAAA
14251 AGCGAAGTTT GAGAAGATAA GGAAGTATGA ACATAACTAA AAAACAAAGT
14301 GGGAAACATT TGTAATACAG AACAGGGCAA TGAAAACCTT GAAGTAAAAT
14351 GGCCATCCCT CAAGAAAGTT CAGGAAATAG TTAACATCAG CTGGGTGCAG
14401 TGGCTCACAC CTATAATCCC AGCACTTTGG AAGGCTGAGG CAGGTGGATC
14451 ACCTGAGGTC AGGAGCTCGA GACCAGTCTG GCCAACATAG TGAAACTCCG
14501 TCTCTGCTAA AAATACAAAA AAAATTAGCC AGGCGTGGTG GTGTGCACCT
14551 GTAATCCAG CTA CTCTGGA GGCTGAGAAG GGAGAATTCC TTGAACCGGG
14601 GAGATGAAGG TTGGAGTGAG CAGAGACCGC GCCATTGCAC TCCAGCCTGG
14651 GCAACAAGAG CGAAGAACAA AACTATGTCT CAAAAAACA AAACACAGCA
14701 AACAAAAATC TATTTTGAAA GAGATGAGAG TGAGCCATAT AACTTGTTTA
14751 AACAAAAGGA AGTTGTGTTG TCGTGTAATT AAATGAAAAT ACTAGGAAGT
14801 GAAATAATAC CTCCAATGGA AATGGTAGAA AGCAGAACTG AAAA ACTTCT
14851 GCTAGGTAGG ATATGGTAGG TCTCTGCACG CCACCACTCC CATTGCAACC
14901 GCTAGGGAAA AAACAGCTAA GATGAAAATG TCTTTTTTTT TCTTTTTTTT
14951 TTTTTTTTGA GATGGAGTCT CGCGCTGTTG CACAGGCTGG AGTGCACTGG
15001 CGCGATCTCA GTTCACTGCA ACCTCTGCCT CTCGGGTTC ACGGATTCTC
15051 CTGCCTCAGC CTCCTGAGTA GCTGGGATTA CAGGCACGCA TCACTCACGA
15101 GCGGCTAATT TTTGTAATTT TAGTAGAGAC GGGGTTTCAA CATGTTGGTC
15151 AGGCTGGTCT CAACTCCTG ACCTCAAAGT GACCCGCCCA CCTCGGCCTC
15201 CCAAAGTGTT GGGATTACAG GGATGAGCCA CCACGCCTGG CCGAAATGTC
15251 TTATTTTAA AAAGAATGAA GAGTGGTCAC AGAAATAAAG ACTGAAT (SEQ ID NO:3)

FEATURES:

Start: 2563
Exon: 2563-2775
Intron: 2776-2927
Exon: 2928-3101
Intron: 3102-4476
Exon: 4477-4573
Intron: 4574-4702
Exon: 4703-4886
Intron: 4887-6827
Exon: 6828-6965
Intron: 6966-7038
Exon: 7039-7158
Intron: 7159-7361
Exon: 7362-7484

FIG. 3G

Intron: 7485-7614
 Exon: 7615-7801
 Intron: 7802-7923
 Exon: 7924-8007
 Intron: 8008-9075
 Exon: 9076-9198
 Intron: 9199-9579
 Exon: 9580-9633
 Intron: 9634-10229
 Exon: 10230-10348
 Intron: 10349-10622
 Exon: 10623-10717
 Intron: 10718-10813
 Exon: 10814-10937
 Intron: 10938-11176
 Exon: 11177-11334
 Intron: 11335-11458
 Exon: 11459-11581
 Intron: 11582-12938
 Exon: 12939-13078
 Stop: 13079

CHROMOSOME MAP POSITION:
 Chromosome 15

ALLELIC VARIANTS (SNPs):

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
28	C	T	Beyond ORF(5')			
847	A	G	Beyond ORF(5')			
2159	G	C	Beyond ORF(5')			
2484	C	T	Beyond ORF(5')			
2577	T	C	Exon	5	S	S
2808	-	T C	Intron			
2922	G	A	Intron			
4312	C	T	Intron			
4903	G	A	Intron			
5193	A	G	Intron			

Context:

DNA

Position

28 CTGGCCACCAGGCTGGCGCAGCCAAGG
 [C,T]
 CGAAGCTCTGGCTGAACCCTGTGCTGGTGTCTGACCACCCTCCCCTCTCTTGACCCGC

FIG. 3H

CTCTCCCGTCAGGGCCCAAGTCCCTGTTTTCTGAGCCCGGGCTGCCTGGGCTGTTGGCAC
TCACAGACCTGGAGCCCCTGGGTGGGTGGTGGGGAGGGGCGCTGGCCCAGCCGGCCTCTC
TGGCCTCCCACCCGATGCTGCTTTCCCTGTGGGGATCTCAGGGGCTGTTTGAGGATATA
TTTTCACTTTGTGATTATTTCACTTTAGATGCTGATGATTTGTTTTGTATTTTAATGG

847 ATGCCGGGGGTTTCATAGGTCACTGGCTCTCCAAGTGCCAGAGGTGGGCAGGTGGTGGCAC
TGAGCCCCCCCCAACACTGTGCCCTGGTGGAGAAAGCACTGACCTGTGATGCCCCCTCAA
ACCTCCTCTTCTGACGTGCCTTTTGACCCCTCCCATTAGGACAATCAGTCCCCTCCCAT
CTGGGAGTCCCCTTTTCTTTTCTACCTAGCCATTCTGGTACCCAGCCATCTGCCAGG
GGTGCCCCCTCCTCTCCCATCCCCCTGCCCTCGTGGCCAGCCCGGCTGGTTTTGTAAGAT
[A,G]
CTGGGTTGGTGACAGTGATTTTTTTCTTGTAAATTAACAGGCCAGCATTGCTGGTTC
TATTTAATGGACATGAGATAATGTTAGAGGTTTTAAAGTGATTAAACGTGCAGACTATGC
AAACCAGGCCAGTCTCCAGTGTGGTACCGTTGCTCCTGCATCGCAGCTGAGGATAGGGG
GCCAGTTAGGCCTACACAGTGGCCTGCCTGCCTGGATGTGGGCCCAAGTCAGAAAGGCCAA
AGTCCTCCAAGGGGCGGGAGGATGCGCCAGCCCCTAGTGGAGGAGCTGGTGCCCCCTGGGG

2159 TGCGGCCAGCGCGGAGCAGGGGGGCTGGCAGGCGTCGGGACGGTCGGGCCGGTCCCCGCC
CGCCCCCTTCCCCTCCACAGGCCCGCCCCGGGGCCTGGGCCAACTGAAACCGCGGGAGGAG
GAAGCGCGGAATCAGGAAGTGGCCGGGTCCGCACCGGGCCTGAGTCGGTCCGAGGCCGT
CCCAGGAGCAGCTGCCCGTGCGGGTACCTCTAGCCCCGGGGCCTGGAGGAGCGGTGGGAG
CTGGGGGCGCGCAGGCAGGGGCAGAGCAGGCGTTCCGAGGGCCAGAGACCCACCCAGGT
[G,C]
GGGGTAGGGGCCGCGGAAGGGCGGGGATGGCCGCAGGGGCAGGGCTCAGGCTGTGGGCGC
CTGAGGCTTCAGCTGGGGCAGGCTTGGCCTGTGAGGACCTGGGCAAGGGTGTCCCTGTA
AGGGGTGGTGGGTGGAAGGGCCTGGGGAGGGAGGCTCCAGGTTGGCTCCTGTTCCCGAAC
GTGCGGAGGAGACCCTGACGCTAAGGAAGCAATGAGGGCCAGTCCCCAGGCCAGGCTGCT
GCTGGGTACCCATGGCTGCGTGTGAGCGAGGCAGGACCCACCTCCTCCCCGTCTGCAGT

2484 GGATGGCCGCAGGGGCAGGGCTCAGGCTGTGGGCGCCTGAGGCTTCAGCTGGGGCAGGCT
TGGCCTGTGAGGACCTGGGCAAGGGTGTCCCTGTAAGGGGTGGTGGGTGGAAGGGCCTG
GGGAGGGAGGCTCCAGGTTGGCTCCTGTTCCCGAACGTGCGGAGGAGACCCTGACGCTAA
GGAAGCAATGAGGGCCAGTCCCCAGGCCAGGCTGCTGCTGGGTACCATGGCTGCGTGTG
AGCGAGGCAGGACCCACCTCCTCCCCGTCTGCAGTCCATCCTGACCTACAGTCCCCAG
[C,T]
CTCCTCGTCCCATGCCTCCGTCTCCAGCTGCTGCCTTGCTCCAGGGATGGCCCCTTTTTC
TGTCGCCAGAACAGCACTATGGGCTTCTCTTCTGAGCTGTGCAGCCCCAGGGCCACGGG
GTCCTGCAGCAATGCAGGAGGCCGAGCTTCGTCTACTGGAGGGCATGAGAAAGTGGATG
GCCCAGCGGGTCAAGAGTGACAGGGAGTATGCAGGACTGCTTCACCACATGTCCCTGCAG
GACAGTGGGGGCCAGAGCCGGGCCATCAGCCCTGACAGCCCCATCAGTCAGGTGGGTCTC

2577 GTAAGGGGTGGTGGGTGGAAGGGCCTGGGGAGGGAGGCTCCAGGTTGGCTCCTGTTCCCG
AACGTGCGGAGGAGACCCTGACGCTAAGGAAGCAATGAGGGCCAGTCCCCAGGCCAGGCT
GCTGCTGGGTACCCATGGCTGCGTGTGAGCGAGGCAGGACCCACCTCCTCCCCGTCTGC
AGTCCATCCTGACCTACAGTCCCCAGCCTCCTCGTCCCATGCCTCCGTCTCCAGCTGCT
GCCTTGCTCCAGGGATGGCCCCTTTTCTGTCCCCAGAACAGCACTATGGGCTTCTCTTC

FIG. 3I

[T,C]
GAGCTGTGCAGCCCCAGGGCCACGGGGTCTGCAGCAAATGCAGGAGGCCGAGCTTCGT
CTACTGGAGGGCATGAGAAAGTGGATGGCCAGCGGGTCAAGAGTGACAGGGAGTATGCA
GGACTGCTTCACCACATGTCCCTGCAGGACAGTGGGGGCCAGAGCCGGGCCATCAGCCCT
GACAGCCCCATCAGTCAGGTGGGTCTCTATGGGACTCTGGTGGGTGCTGGCGTATCTGCC
TTCTCCTTCTCTCCTGGGGGCCCTCTGGGGCAGTGGCTGGAGATCTGGCAGGCCAATGC

2808
CCAGCTGCTGCCTTGCCTCCAGGGATGGCCCTTTTCTGTCCCCAGAACAGCACTATGGG
CTTCTCTTCTGAGCTGTGCAGCCCCAGGGCCACGGGGTCTGCAGCAAATGCAGGAGGC
CGAGCTTCGTCTACTGGAGGGCATGAGAAAGTGGATGGCCAGCGGGTCAAGAGTGACAG
GGAGTATGCAGGACTGCTTCACCACATGTCCCTGCAGGACAGTGGGGGCCAGAGCCGGGC
CATCAGCCCTGACAGCCCCATCAGTCAGGTGGGTCTCTATGGGACTCTGGTGGGTGCTGG
[-,T,C]
GTATCTGCCTTCTCCTTCTCCTGGGGGCCCTCTGGGGCAGTGGCTGGAGATCTGGCA
GGCCAATGCTTGGGAGCCATTGTGCCCCCTCCCTGCCTCCCCATCTGTGCTGTATAGT
CCTGGGCTGAGATCACCAGCCAACTGAGGGCCTGAGCCGCTTGCTGCGGCAGCACGCAG
AGGATCTGAACTCAGGGCCCCCTGAGCAAGCTGAGCCTGCTCATCCGGGAACGGCAGCAGC
TTCGAAGACCTACAGCGAGCAGTGGCAGCAGCTGCAGCAGGAGCTCACCAAGGTGAGCG

2922
GGAGGCCGAGCTTCGTCTACTGGAGGGCATGAGAAAGTGGATGGCCAGCGGGTCAAGAG
TGACAGGGAGTATGCAGGACTGCTTCACCACATGTCCCTGCAGGACAGTGGGGGCCAGAG
CCGGGCCATCAGCCCTGACAGCCCCATCAGTCAGGTGGGTCTCTATGGGACTCTGGTGGG
TGCTGGCGTATCTGCCTTCTCCTTCTCCTGGGGGCCCTCTGGGGCAGTGGCTGGAGA
TCTGGCAGGCCAATGCTTGGGAGCCATTGTGCCCCCTCCCTGCCTCCCCATCTGTGCT
[G,A]
TATAGTCCTGGGCTGAGATCACCAGCCAACTGAGGGCCTGAGCCGCTTGCTGCGGCAGC
ACGCAGAGGATCTGAACTCAGGGCCCCCTGAGCAAGCTGAGCCTGCTCATCCGGGAACGGC
AGCAGCTTCGCAAGACCTACAGCGAGCAGTGGCAGCAGCTGCAGCAGGAGCTCACCAAGG
TGAGCGGGCAGCACTGGGGCTTCGGTCATTTCTGTCTAAATTTTGAGCCTCGAAGGGGT
GTTTTGCACAAGAGGCCCTGGATTCACTGGGGAAGTGTAAATCCCTGACCGCAGGCCTGG

4312
GGACGAGAACACTGATCTCGGGCTGTCTATCTATTCTACTGCCAAGAACATAATTTGCAG
GACCCAGTGCAAAGTGAAATTGTGGGGTCTTTGTTAAAAGATTGCTAGGAATTTCCAGG
TGGCAATAATGGAGAATGAAACCAAGCACAGGGCCCTTCTACATGTGGAGCCCCGTGTGA
CTGCACAGGCCGTGCACACCTGCAACTGGCCCTGCCTGCCACCAGGCTACCACTGTCAGT
CCAAGGAGGGACCGTTGTAGCCTGTAGTCTACCTCTTTGCCTCCCCAAGGGGTCTGTCTT
[C,T]
AACAGGCTCTCTGATCTTTGACTCTCACGTGAGCAGCCAGCTTTCCAGAAAGTCTCCAGG
TGCTCCTTGCCTGACGACAGGACCTTTCCAGGGCTTCACCCCAGGCAAGAATCTTCACA
ACTGGGGACCTGCTGCCCCCACTGGCCTCTCCTCTCTCCCTAGACCCACAGCCAGGACA
TTGAGAAGCTGAAGAGCCAGTACCGAGCTCTGGCACGGGACAGTGCCCAAGCCAAGCGCA
AGTACCAGGAGGCCAGCAAAGGTTCTGTGGCTTCCCTTGCTGGCAGGGAGGGAATCCGAAG

4903
GAATCCGAAGCCAGTGCTGACCTGTCCTTGGGTACCCAGAGAGTGGGGGCTGCCTGGGCC
TCCATGCTGTCTATACCCCTTGCCCCCTTCTGGCAGACAAGGACCGTGACAAGGCC
AAGGACAAGTATGTGCGCAGCCTGTGAAGCTCTTTGCTCACCACAACCGCTATGTGCTG

FIG. 3J

GGCGTGCGGGCTGCGCAGCTACACCACCAGCACCACCAGCTCCTGCTGCCCCGGCCTG
CTGCGGTCACTGCAGGACCTGCACGAGGAGATGGCTTGCATCCTGTAAGCCCGCAGCCCC
[G, A]

TCCCCTGGCCCCCACCCTTGAGCAGCCCTAAGCCCAGCCATCAGGCCCAGAGGCAGGACC
CAGAAAATCCATTGCTGGGAAGGTGCTGGCCATGTAACCACATGAGAACGGGACCTGGGC
CAAGGATTGGAAACAGGCAACTTACCTCTGAATTACACTATTCCAGGGTCTCATTATTCC
AGGGTTTTATTACATTATTGAGCACTGTTCTGGGCTCTGGATTATACCAGAGAACGATG
GTAGACAAAAACATCTGTCCTCAGGGATCTTTCGTGTTAGTGGAGTGAGAATGTGAGGAG

5193

CCGCAGCCCCGTCCCCTGGCCCCCACCCTTGAGCAGCCCTAAGCCCAGCCATCAGGCCCCA
GAGGCAGGACCCAGAAAATCCATTGCTGGGAAGGTGCTGGCCATGTAACCACATGAGAAC
GGGACCTGGGCCAAGGATTGGAAACAGGCAACTTACCTCTGAATTACACTATTCCAGGGT
CTCATTATTCCAGGGTTTTATTACATTATTGAGCACTGTTCTGGGCTCTGGATTATACC
AGAGAACGATGGTAGACAAAAACATCTGTCCTCAGGGATCTTTCGTGTTAGTGGAGTGAG
[A, G]

ATGTGAGGAGCACTAAGAGCCATGGAGAAAAATAAAGCAAGAGAAGTGGATCGGGACCTG
GGAGCACGGAGGCAAGGGAGGAGGTGACAGTTGTCCATAGAGTGATCTGGGAAAGCCTCT
TGAGAGGTGACATTCAAAGAGGCCCTGAGAGGGGTACGGGAGTGAATCATGGGGCTATT
TGGAGAAAGACCATTCAGAAAGGAGGACAGCAATTACACAGGCCTTGAGGTAGGAGAGT
ACCAGGGACTAATAGCCAGGAACCAAGTGGTGCCTCTGAGAGTGAGGGAGGGGAGAGTCA